

# Particle-Hole Asymmetry in Gapped Topological Insulator Surface States

C. J. Tabert<sup>1,2</sup> and J. P. Carbotte<sup>3,4</sup>

<sup>1</sup>*Department of Physics, University of Guelph, Guelph, Ontario N1G 2W1 Canada*

<sup>2</sup>*Guelph-Waterloo Physics Institute, University of Guelph, Guelph, Ontario N1G 2W1 Canada*

<sup>3</sup>*Department of Physics, McMaster University, Hamilton, Ontario L8S 4M1 Canada and*

<sup>4</sup>*Canadian Institute for Advanced Research, Toronto, Ontario M5G 1Z8 Canada*

(Dated: May 26, 2015)

We consider the combined effect of a gap and a Zeeman interaction on the helical Dirac fermions which exist on the surface of a topological insulator. Magneto-optical properties, the magnetization, Hall effect and the density of states are considered with emphasis on the particle-hole asymmetry which arises when a subdominant Schrödinger piece is included along with the dominant Dirac part of the Hamiltonian. When appropriate, we compare our results with those of a single valley gapped graphene system for which Zeeman splitting behaves differently. We provide a derivation of the phase offset in the magnetic oscillations brought about by the combined effect of the gap and Schrödinger term without requiring the semiclassical Onsager quantization condition. Our results agree with previous discussions based on semiclassical arguments.

PACS numbers: 75.60.Ej, 72.80.Vp, 71.70.Di, 73.43.-f

## I. INTRODUCTION

A topological insulator (TI)<sup>1,2</sup> is a material which is insulating in the bulk but possesses a metallic spectrum of surface states. These are expected to have high mobility as they are protected by the topology of the band structure. The surface states of Bi<sub>2</sub>Se<sub>3</sub> have a single Dirac cone<sup>3</sup> centred at the  $\Gamma$ -point of the two-dimensional (2D) surface Brillouin zone. Other materials can have a different odd number of such points. As an example, samarium hexaboride has three<sup>4</sup>. Such a new phase of matter comes about through a band inversion caused by a strong spin-orbit coupling. At present, these systems are extensively studied due to their novel physics and their possible applications in devices such as quantum computing<sup>5</sup> and photonics<sup>6</sup>. An important observation is the phenomenon of spin-momentum locking<sup>3</sup> established by spin- and angular resolved photoemission spectroscopy (ARPES). This shows that the in-plane component of the spin is perpendicular to the in-plane momentum. A gap can also be introduced at the Dirac point through doping of the TI surface with magnetic (time-reversal-breaking) impurities. In the work of Chen et al.<sup>7</sup>, a gap of  $\Delta \sim 7$  meV was observed through ARPES measurements on a sample of (Bi<sub>0.99</sub>Mn<sub>0.01</sub>)<sub>2</sub>Se<sub>3</sub> with 1% Mn substitution on the Bi site. Another avenue to create a gap is to make the TI ultra thin with a distance between the top and bottom surface of the order of the extent in space (perpendicular to their plane) of the surface states<sup>8–10</sup>. Tunnelling between the top and bottom surfaces gaps the electronic spectrum. Of course, in this case, both surfaces can contribute to a particular phenomenon and a sum over both cones (which have gaps of opposite sign) is needed as has been discussed recently by Yoshimi et al.<sup>11</sup> and Zhang et al.<sup>12</sup> in the context of the quantum Hall effect. Neupane et al.<sup>13</sup> find that varying the quantum-tunnelling gap in ultra-thin films leads to a modulation of the in-plane spin texture. While the spin-

momentum locking remains, the in-plane component of spin can itself be decreased with decreasing thickness. Other such studies include work by Tahir et al.<sup>14</sup> on the oscillations expected in the quantum capacitance of thin films. For simplicity, we will consider a single gapped Dirac cone; however, the results applicable to thin films can be obtained by assembling two such cones with the appropriate relationship between their respective gaps.

Particle-hole asymmetry is an important feature of the surface-fermion band structure of a TI. This arises from a subdominant quadratic-in-momentum Schrödinger term with mass  $m$  in addition to the dominant linear-in-momentum Dirac term with Fermi velocity  $v_F$ . The relativistic piece has its origin in the strong spin-orbit coupling and involves the Pauli spin-matrices  $\hat{\sigma}_x$  and  $\hat{\sigma}_y$ . This is distinct from the case of graphene where the matrices exist in pseudospin space and the energy levels are degenerate in spin<sup>15</sup>. With the application of a magnetic field perpendicular to the plane, the Landau levels (LLs) are split by the Zeeman field and form two sublevels with a definite  $s_z$  component of spin  $\pm \hbar/2$ <sup>15</sup>. This contrasts with the case of the TI for which only the  $N = 0$  level is fully spin polarized in  $s_z$ <sup>16</sup> while the average  $s_z$  for all the other levels is small in comparison. For the presently studied TI, the Dirac magnetic energy scale at one Tesla is an order of magnitude larger than the Schrödinger scale. However, at  $B = 36$  T (for example) the two scales differ by less than a factor of two. Even when  $B$  is small, the Schrödinger term can lead to novel effects such as the splitting of the optical absorption lines seen in the real part of the AC conductivity<sup>17</sup>. Thus, this term must be included in any complete theoretical treatment and is important when considering experimental data. In particular, optical data gives valuable information on the dynamics of the surface charge carriers in a TI as demonstrated in the experimental work by LaForge et al.<sup>18</sup> and Orlita et al.<sup>19</sup> among others.

Our paper is organized as follows: in Sec. II, we specify

our low-energy Hamiltonian involving Schrödinger-Dirac kinetic energy and the gap term  $\Delta\hat{\sigma}_z$  which does not commute with the rest of the Hamiltonian. The eigenenergies and eigenfunctions which arise in the presence of a magnetic field are worked out; from these, the average value of  $s_z$  for each level is determined. We also consider a related system written in pseudospin space. In this case, the wave-functions are simultaneous eigenstates of the Hamiltonian and  $\hat{s}_z$ . Thus, each LL is fully spin polarized in the  $z$  direction. In Sec. III, we begin our discussion by considering the integrated density of states up to energy  $\omega_{\max}$  (as measured from the Dirac point). We emphasize the effect of particle-hole asymmetry in a TI and contrast this with the case of a single valley of a graphene-like system which shows no asymmetry in the absence of a gap. We also discuss the modifications brought about by a finite  $\Delta$  and by Zeeman splitting. Section IV contains formulas for the magneto-optical conductivity. Both the longitudinal and the transverse Hall conductivities are considered. Again, particle-hole asymmetry is emphasized. We find a comparison with the single-valley graphene-like system to be helpful. Circular dichroism is discussed. In Sec. V, we examine the magnetization of the metallic surface states. Through the Streda relation<sup>20,21</sup>, we relate its derivative with respect to chemical potential ( $\mu$ ) to the quantization of the Hall plateaus. As expected, we find complete agreement between this and our conductivity formula. We find that the quantization remains half-integral (i.e.  $\pm 1/2, \pm 3/2, \pm 5/2, \dots$ ) which is characteristic of relativistic fermions (with due consideration for the degeneracy factors). The value of  $\mu$  at which a new step occurs is changed with the value of Schrödinger mass,  $\Delta$  and Zeeman splitting. In contrast to graphene, Zeeman interactions do not introduce additional steps<sup>15</sup> between those present when the effect is neglected. It is important to realize that when the Dirac term in our Hamiltonian is dominant, this describes a TI and the Hall quantization is relativistic. When the Schrödinger piece dominates, the Hall quantization is non-relativistic and the Hamiltonian applies to the usual spintronic materials. In Sec. VI, we consider the magnetic oscillations and in particular, emphasize the phase offset in the usual semiclassical expression. This quantity is  $1/2$  for non-relativistic Schrödinger particles and  $0$  for relativistic Dirac fermions. In the presence of a finite mass and magnetic gap, we find that the offset is changed to  $\gamma = -\Delta(1+g)/(2mv_F^2)$ , where  $g$  is the Zeeman strength. This reduces to the relativistic result for either  $\Delta = 0$  or  $m \rightarrow \infty$ . The phase offset depends on the Zeeman term through the coupling coefficient  $g$ . Except for this factor, our quantum mechanical results agree with a previous result<sup>22</sup> which relied on semiclassical arguments<sup>23-29</sup>. Fuchs et al.<sup>24</sup> start with the Bohr-Sommerfeld quantization condition suggested by Onsager<sup>30</sup> and include a correction in the band energies due to the magnetization  $\mathcal{M}(\mathbf{k})$  induced by the external magnetic field  $B$  of the form  $-\mathcal{M}(\mathbf{k}) \cdot \mathbf{B}$ . They apply this to a two band model of gapped graphene and explic-

itly show that the phase offset ( $\gamma$ ) in the magnetic oscillations is zero in this case even though the Berry phase around the cyclotron orbit is not  $\pi$ . Because of the gap, it is instead equal to  $\pi W_c[1 - \Delta/\mu]$ <sup>31</sup> where  $W_c$  is the winding number and is a topological invariant referred to by Fuchs et al.<sup>24</sup> as the topological part of the Berry phase. They establish that  $\gamma = 1/2 - W_c/2 = 0$  in this case and it is only the topological part of the Berry phase which enters this quantity. Thus, the  $-W_c/2$  exactly cancels the Maslov index contribution of  $1/2$ <sup>24</sup>. Following this line of reasoning, Wright and McKenzie<sup>22</sup> extend this semiclassical method to the case when a Schrödinger mass is introduced as well as a gap. In this case, the result for the phase offset is given by their Eqn. (26). To lowest order in the Schrödinger mass, this reduces to  $\Delta/(2mv_F^2)$ . We derive this result without resorting to any semiclassical arguments, and further find its generalization to include the Zeeman term.

## II. SURFACE STATE HAMILTONIAN

In the presence of magnetic dopants, the helical surface states of a TI are given by the Bychkov-Rashba Hamiltonian<sup>32,33</sup>

$$\hat{H} = \frac{\hbar^2 k^2}{2m} + \hbar v_F(k_x \hat{\sigma}_y - k_y \hat{\sigma}_x) + \Delta \hat{\sigma}_z, \quad (1)$$

where  $\hat{\sigma}_x$ ,  $\hat{\sigma}_y$  and  $\hat{\sigma}_z$  are the usual Pauli-spin matrices and  $\hbar \mathbf{k}$  is the momentum measured relative to the  $\Gamma$ -point of the surface Brillouin zone. The first term is the Schrödinger piece for describing electrons with effective mass  $m$ . The linear-in-momentum term describes massless Dirac fermions which move with a Fermi velocity  $v_F$ . The last term accounts for the magnetic impurities which open a gap of  $2\Delta$  in the band structure. For  $\text{Bi}_2\text{Te}_3$ , band structure calculations<sup>34,35</sup> predict:  $v_F = 4.3 \times 10^5 \text{ m/s}$ , and  $m = 0.09m_e$  where  $m_e$  is the bare mass of an electron<sup>36</sup>. Solving Eqn. (1) yields the energy dispersion

$$E_{\pm}(\mathbf{k}) = \frac{\hbar^2 k^2}{2m} \pm \sqrt{(\hbar v_F k)^2 + \Delta^2}. \quad (2)$$

A schematic plot of the surface-state band structure is given in Fig. 1. Figure 1(a) shows the gapless spectrum which we contrast to frame (b) where a finite  $\Delta$  is included. It is important to note that we are mapping a continuum Hamiltonian onto a finite Brillouin zone. Equation 1 permits the valence band to bend back toward the zero energy axis; this is unphysical for TIs. As a result, an appropriate momentum cutoff must be applied.

To examine the magnetic response of such systems<sup>17,37-47</sup>, we must consider the effect of a finite magnetic field  $B$  oriented perpendicular to the surface ( $\hat{z}$ ). In the Landau gauge, the magnetic vector potential, defined through  $\mathbf{B} = \nabla \times \mathbf{A}$ , may be written as  $\mathbf{A} = (0, Bx, 0)$ . We account for the magnetic field by

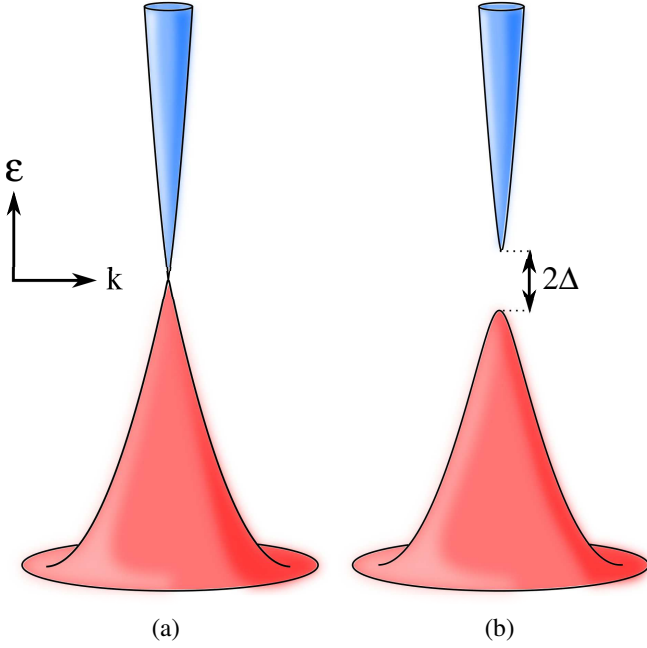


FIG. 1. (Color online) (a) Gapless spectrum of topological surface states. (b) The effect of including magnetic dopants on the low-energy band structure. In both cases, significant particle-hole asymmetry is observed as the mass term causes the conduction band (blue) to narrow while the valence band (red) fans out.

applying the usual Peierls substitution to the momentum ( $\hbar\mathbf{k} \rightarrow \hbar\mathbf{k} + e\mathbf{A}$ ) in Eqn. (1). Our Hamiltonian becomes

$$\hat{H} = \frac{\hbar^2}{2m} [(-i\partial_x)^2 + (-i\partial_y + eBx/\hbar)^2] + \Delta\hat{\sigma}_z + \hbar v_F [(-i\partial_x)\hat{\sigma}_y - (-i\partial_y + eBx/\hbar)\hat{\sigma}_x]. \quad (3)$$

A Zeeman interaction can be included by adding  $-(1/2)g_s\mu_B B\hat{\sigma}_z$ , where  $g_s$  is the coupling strength ( $\approx 8$  for  $\text{Bi}_2\text{Se}_3$ <sup>21</sup>) and  $\mu_B = e\hbar/(2m_e) \approx 5.78 \times 10^{-2}$  meV/T is the Bohr magneton. Next, we define the raising and lowering operators

$$a^\dagger \equiv \frac{l_B}{\sqrt{2}} \left[ -ik_x + \frac{x + x_0}{l_B^2} \right], \quad (4)$$

and

$$a \equiv \frac{l_B}{\sqrt{2}} \left[ ik_x + \frac{x + x_0}{l_B^2} \right], \quad (5)$$

where  $l_B \equiv \sqrt{\hbar/(eB)}$  and  $x_0 \equiv k_y l_B^2$ . On a Fock state ( $|N\rangle$ ) of the harmonic oscillator Hamiltonian, these operators have the property:

$$a|N\rangle = \sqrt{N}|N-1\rangle, \quad (6)$$

and

$$a^\dagger|N\rangle = \sqrt{N+1}|N+1\rangle. \quad (7)$$

With these definitions, the Hamiltonian reduces to

$$\hat{H} = \begin{pmatrix} \frac{\hbar^2}{ml_B^2} \left[ a^\dagger a + \frac{1}{2} \right] + \mathcal{Z} & -\hbar v_F \frac{\sqrt{2}}{l_B} a \\ -\hbar v_F \frac{\sqrt{2}}{l_B} a^\dagger & \frac{\hbar^2}{ml_B^2} \left[ a^\dagger a + \frac{1}{2} \right] - \mathcal{Z} \end{pmatrix}, \quad (8)$$

where  $\mathcal{Z} \equiv \Delta - g_s\mu_B B/2$ . Equation (8) yields the LL dispersion

$$\mathcal{E}_{N,s} = \begin{cases} E_0 N + s \sqrt{2NE_1^2 + \left[ \Delta - \frac{E_0}{2}(1+g) \right]^2} & N = 1, 2, 3, \dots \\ \frac{E_0}{2}(1+g) - \Delta & N = 0 \end{cases}, \quad (9)$$

where  $E_1 \equiv \hbar v_F \sqrt{eB/\hbar}$ ,  $E_0 \equiv \hbar eB/m$ ,  $g = g_s m/(2m_e)$  is the renormalized Zeeman coupling coefficient, and  $s = \pm$  for the conduction and valence band, respectively. It is important to note that in a sum over  $s$ , the  $N = 0$  level will only contribute once. Unless otherwise noted, we will take  $E_1/\sqrt{B} = 10.4$  meV/ $\sqrt{\text{T}}$ ,  $E_0/B = 1.1$  meV/T and  $B = 1$  T. We note that Eqn. (9) is identical to that given previously by Shen et al.<sup>48,49</sup>. However, these authors are only interested in the limit of spintronic materials for which the Schrödinger term in Eqn. (1) is dominant. Here, we treat the opposite limit where the Dirac term is dominant (requiring a band cutoff). Thus, there is no overlap between the two studies and the works are complimentary.

Since our Hamiltonian is written in the spin basis, the eigenstates are comprised of spin-up and -down amplitudes. Indeed, the wave-functions of Eqn. (8) are

$$|Ns\rangle = \begin{pmatrix} C_{N,s}^\uparrow |N-1\rangle \\ C_{N,s}^\downarrow |N\rangle \end{pmatrix}, \quad (10)$$

where

$$C_{N,s}^\uparrow = \begin{cases} -s \sqrt{\frac{1}{2} + s \frac{\Delta - (1+g)E_0/2}{2(\mathcal{E}_{N,+} - E_0 N)}} & N = 1, 2, 3, \dots \\ 0 & N = 0 \end{cases}, \quad (11)$$

is the spin-up component and

$$C_{N,s}^\downarrow = \begin{cases} \sqrt{\frac{1}{2} - s \frac{\Delta - (1+g)E_0/2}{2(\mathcal{E}_{N,+} - E_0 N)}} & N = 1, 2, 3, \dots \\ 1 & N = 0 \end{cases}, \quad (12)$$

gives the spin-down amplitude. We immediately see that the  $N = 0$  LL is entirely populated by spin-down electrons<sup>16</sup> (i.e.  $C_{N,s}^\uparrow = 0$ ). To see the  $z$  component of spin of the remaining levels, we compute the average

$$\langle \hat{s}_z \rangle = \langle N | \frac{\hbar}{2} \hat{\sigma}_z | N \rangle. \quad (13)$$

This gives

$$\langle \hat{s}_z \rangle = \frac{\hbar}{2} \begin{cases} \frac{s [\Delta - (1+g)E_0/2]}{\sqrt{2E_1^2 N + [\Delta - (1+g)E_0/2]^2}} & N \neq 0 \\ -1 & N = 0 \end{cases}. \quad (14)$$

Clearly, an  $\hat{s}_z$  spin polarization is quickly lost for  $N \neq 0$ .

Examining the  $N = 0$  level [Eqn. (9)] reveals that for  $\Delta = 0$ , the LL sits at positive energy (i.e.  $\mathcal{E}_{0,+} = (1+g)E_0/2$ ). For finite  $\Delta$ , the level remains above zero as long as  $(1+g)E_0/2 > \Delta$ . For  $\Delta > (1+g)E_0/2$ , the zeroth level is situated at negative energy. Also, in contrast to graphene (where spin is a good quantum number), a finite Zeeman term does not split the levels but simply renormalizes their energy. As opposed to graphene, the  $N \neq 0$  levels are not simply shifted by  $\pm g_s \mu_B B/2$ . The zeroth level is translated in the usual way since it has a definite spin-down polarization. A schematic plot of the low-energy LLs in a TI is shown in Fig. 2. Particular

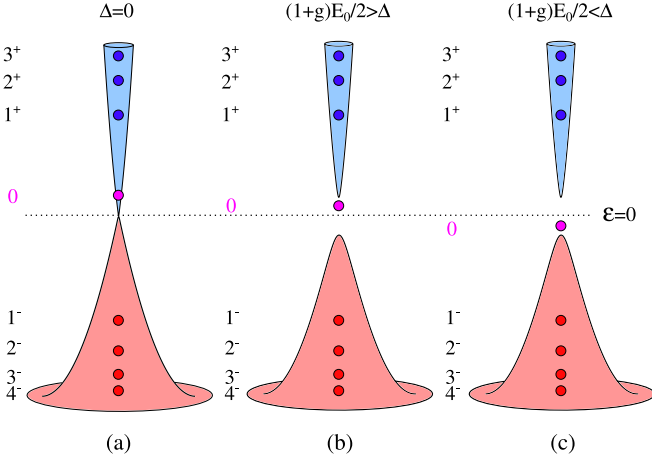


FIG. 2. (Color online) Schematic illustration of the Landau level energies for a TI with (a)  $\Delta = 0$ , (b)  $(1+g)E_0/2 > \Delta$  and (c)  $\Delta > (1+g)E_0/2$ . As  $\Delta$  increases, the zeroth level begins at positive energy and decreases (becoming negative when  $\Delta > (1+g)E_0/2$ ).

attention is given to the relative location of the  $N = 0$  level. The various LL energies are given by the coloured circles (blue for conduction levels, red for valence levels, and purple for  $N = 0$ ) and are overlaid on the  $B = 0$  band structure.

Throughout this paper, we will be interested in comparing the results of a TI with the familiar case of gapped graphene; in particular, the graphene result at the  $K$ -point. In the presence of an asymmetry gap  $\Delta$ , the graphene Hamiltonian written about a single valley  $\xi$  in the sublattice basis becomes

$$\hat{H}_\xi = \begin{pmatrix} \Delta & \hbar v(\xi k_x - i k_y) \\ \hbar v(\xi k_x + i k_y) & -\Delta \end{pmatrix}, \quad (15)$$

where  $\xi = \pm$  for the  $K$  and  $K'$  points of the hexagonal Brillouin zone. Applying the same techniques as before, the LL energies are<sup>37–39</sup>

$$\mathcal{E}_{N,s}^{\xi\sigma} = \begin{cases} -\frac{1}{2}g_s\mu_B B\sigma + s\sqrt{\Delta^2 + 2NE_1^2}, & N = 1, 2, 3, \dots \\ -\xi\Delta - \frac{1}{2}g_s\mu_B B\sigma, & N = 0 \end{cases}, \quad (16)$$

where  $\sigma = \pm$  for spin up and down, respectively. The corresponding wave-functions are

$$|Ns\rangle_K = \begin{pmatrix} -i\mathcal{A}_{N,s}|N-1\rangle \\ \mathcal{B}_{N,s}|N\rangle \end{pmatrix} \quad (17)$$

and

$$|Ns\rangle_{K'} = \begin{pmatrix} -i\mathcal{A}_{N,s}|N\rangle \\ \mathcal{B}_{N,s}|N-1\rangle \end{pmatrix}, \quad (18)$$

respectively, where<sup>37–39</sup>

$$\mathcal{A}_{N,s} = \begin{cases} \frac{s\sqrt{|\mathcal{E}_{N,s}^{\xi\sigma}| + s\Delta}}{\sqrt{2|\mathcal{E}_{N,s}^{\xi\sigma}|}}, & N \neq 0, \\ \frac{1-\xi}{2}, & N = 0, \end{cases} \quad (19)$$

and

$$\mathcal{B}_{N,s} = \begin{cases} \frac{\sqrt{|\mathcal{E}_{N,s}^{\xi\sigma}| - s\Delta}}{\sqrt{2|\mathcal{E}_{N,s}^{\xi\sigma}|}}, & N \neq 0, \\ \frac{1+\xi}{2}, & N = 0. \end{cases} \quad (20)$$

Unlike the TI, spin is a good quantum number and thus a Zeeman interaction simply shifts the spin-up (-down) levels down (up) by a constant amount.

### III. DENSITY OF STATES

We begin our analysis with a comparison of the integrated density of states. In a magnetic field, the density of states is given by a sum of  $\delta$ -functions which peak at the various LL energies. For a TI,

$$N(\omega) = \frac{eB}{h} \left[ \delta(\omega - \mathcal{E}_{0,+}) + \sum_{\substack{N=1 \\ s=\pm}}^{\infty} \delta(\omega - \mathcal{E}_{N,s}) \right], \quad (21)$$

where  $\mathcal{E}_{N,s}$  is given by Eqn. (9). Similarly, for graphene

$$N_{\xi\sigma}(\omega) = \frac{eB}{h} \left[ \delta(\omega - \mathcal{E}_{0,+}^{\xi\sigma}) + \sum_{\substack{N=1 \\ s=\pm}}^{\infty} \delta(\omega - \mathcal{E}_{N,s}^{\xi\sigma}) \right], \quad (22)$$

where  $\mathcal{E}_{N,s}^{\xi\sigma}$  is given by Eqn. (16). To compare the two results, we define the total integrated density of states up to energy  $\omega_{\max}$  as

$$I(\omega_{\max}) = \int_0^{\omega_{\max}} N(\omega) d\omega. \quad (23)$$

A plot of  $I(\omega_{\max})$  as function of the cutoff energy  $\omega_{\max}$  is shown in Fig. 3. For graphene, the steps are of double weight due to spin-degeneracy. Figure 3(a) shows the

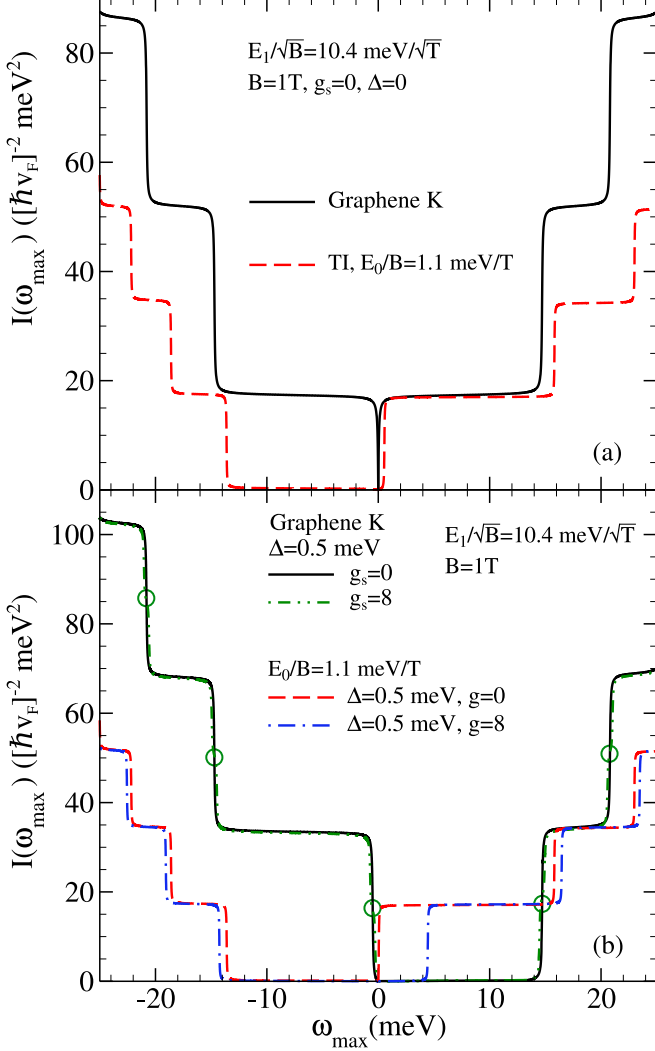


FIG. 3. (Color online) Integrated density of states for graphene at the  $K$ -point and a TI (a) without and (b) with a gap of  $\Delta = 0.5 \text{ meV}$ . (a) For  $\Delta = 0$ , particle-hole symmetry is present in graphene. For a TI the valence and conduction levels are asymmetric. (b) For finite  $\Delta$ , the symmetry in graphene is removed. A Zeeman interaction spin-splits the graphene levels and simply renormalized those of a TI.

$\Delta = 0$  results for graphene at the  $K$ -point (solid black) and a TI (dashed red) in the absence of Zeeman splitting. For graphene, the  $N = 0$  LL is pinned at zero energy and particle-hole symmetry is present for all  $N$ .

For the TI, the  $N = 0$  level sits at positive energy and clear particle-hole asymmetry exists for the  $N \neq 0$  levels. The valence levels sit closer to  $\omega_{\max} = 0$  than the corresponding conduction levels. In Fig. 3(b), a finite gap is included and the effect of Zeeman splitting is explored. For graphene with  $g_s = 0$  (solid black), the gap moves the  $N = 0$  level to negative energy at the  $K$ -point and breaks the particle-hole symmetry. With a finite Zeeman term (dash-double-dotted green), the steps split into two (emphasized by the green circles). This is characteristic of systems with spin-degeneracy. For a TI (dashed red), the gap shifts the energy levels and a finite Zeeman interaction (dash-dotted blue) does not split the steps but further renormalizes their energy. A plot of the LLs for  $\Delta = 0$  and varying  $g_s$  is shown in Fig. 4. For graphene

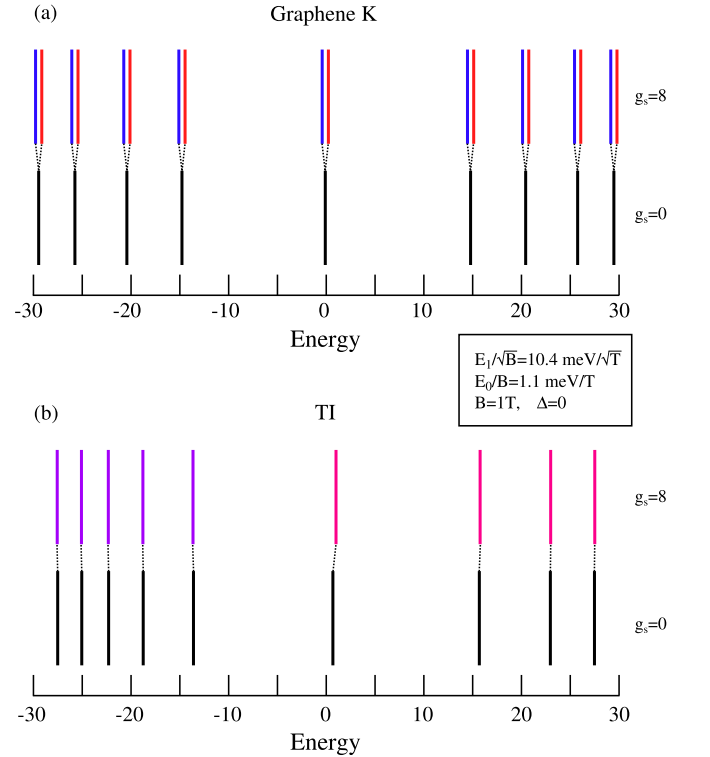


FIG. 4. (Color online)  $\Delta = 0$  LLs in (a) graphene and (b) a TI with and without Zeeman splitting. (a) In graphene, a Zeeman interaction spin-splits all the LLs and shifts the spin-down levels (red) up while the spin-up LLs (blue) decrease in energy. (b) For a TI, Zeeman effects do not split the levels.

at  $K$  [frame (a)], the  $g_s = 0$  levels are symmetric. A finite Zeeman splitting causes the spin-down levels (red) to increase in energy by  $g_s \mu_B B/2$  while the spin-up levels (blue) decrease by the same amount. This splitting does not depend on  $N$  and creates a series of spin-polarized levels. The TI [frame (b)] is particle-hole asymmetric and a Zeeman term causes the conduction levels (pink) to increase in energy while the valence levels (purple) decrease. Unlike graphene, the  $N \neq 0$  levels are not spin-polarized and the splitting is not uniform; for larger

$N$ , the renormalization becomes less prominent.

#### IV. MAGNETO-OPTICAL CONDUCTIVITY

In the one-loop approximation, the optical conductivity  $\sigma_{\alpha\beta}(\Omega)$  is given by the familiar Kubo formula

$$\sigma_{\alpha\beta}(\Omega) = \frac{i\hbar}{2\pi l_B^2} \sum_{\substack{N,M=0 \\ s,s'=\pm}}^{\infty} \frac{f_{M,s'} - f_{N,s}}{\mathcal{E}_{N,s} - \mathcal{E}_{M,s'}} \times \frac{\langle \bar{N}s | \hat{j}_\alpha | \bar{M}s' \rangle \langle \bar{M}s' | \hat{j}_\beta | \bar{N}s \rangle}{\hbar\Omega + \mathcal{E}_{M,s'} - \mathcal{E}_{N,s} + i\hbar/(2\tau)}, \quad (24)$$

where  $f_{n,s}$  is the Fermi function for state  $n$  in band  $s$ ,  $\tau$  is the relaxation time and  $\hat{j}_\alpha = e\hat{v}_\alpha \equiv (e/\hbar)(\partial\hat{H}/\partial k_\alpha)$  is the current operator. In the zero temperature limit,  $f_n$  can be replaced by the Heaviside step function  $\Theta(\mu - \mathcal{E}_{n,s})$ . The necessary velocity operators are

$$\hat{v}_x = \frac{\hbar}{m}k_x + v_F\hat{\sigma}_y = i\frac{\hbar}{m}\frac{a^\dagger - a}{\sqrt{2}l_B} + v_F\hat{\sigma}_y, \quad (25)$$

and

$$\hat{v}_y = \frac{\hbar}{m}\left[k_y + \frac{eBx}{\hbar}\right] - v_F\hat{\sigma}_x = \frac{\hbar}{m}\frac{a^\dagger + a}{\sqrt{2}l_B} - v_F\hat{\sigma}_x. \quad (26)$$

Calculating the appropriate matrix elements, we obtain the real and imaginary parts of the longitudinal conductivity

$$\text{Re} \left\{ \frac{\sigma_{xx}(\Omega)}{e^2/\hbar} \right\} = \frac{E_1^2}{2\pi} \sum_{\substack{N,M=0 \\ s,s'=\pm}}^{\infty} \frac{f_{M,s'} - f_{N,s}}{\mathcal{E}_{N,s} - \mathcal{E}_{M,s'}} \times \frac{\Gamma}{(\Omega + \mathcal{E}_{M,s'} - \mathcal{E}_{N,s})^2 + \Gamma^2} \times [\mathcal{F}(Ns; Ms')\delta_{N,M-1} + \mathcal{F}(Ms'; Ns)\delta_{M,N-1}], \quad (27)$$

and

$$\text{Im} \left\{ \frac{\sigma_{xx}(\Omega)}{e^2/\hbar} \right\} = \frac{E_1^2}{2\pi} \sum_{\substack{N,M=0 \\ s,s'=\pm}}^{\infty} \frac{f_{M,s'} - f_{N,s}}{\mathcal{E}_{N,s} - \mathcal{E}_{M,s'}} \times \frac{\Omega + \mathcal{E}_{M,s'} - \mathcal{E}_{N,s}}{(\Omega + \mathcal{E}_{M,s'} - \mathcal{E}_{N,s})^2 + \Gamma^2} \times [\mathcal{F}(Ns; Ms')\delta_{N,M-1} + \mathcal{F}(Ms'; Ns)\delta_{M,N-1}], \quad (28)$$

respectively, where  $\Gamma \equiv \hbar/(2\tau)$  and

$$\mathcal{F}(Ns; Ms') \equiv \left[ c_{M,s'}^\dagger c_{N,s}^\dagger - \frac{E_0}{\sqrt{2}E_1} \right] \times \left( \sqrt{N}c_{M,s'}^\dagger c_{N,s}^\dagger + \sqrt{N+1}c_{M,s'}^\dagger c_{N,s}^\dagger \right)^2. \quad (29)$$

Likewise, the real and imaginary parts of the transverse Hall conductivity are

$$\text{Re} \left\{ \frac{\sigma_{xy}(\Omega)}{e^2/\hbar} \right\} = \frac{E_1^2}{2\pi} \sum_{\substack{N,M=0 \\ s,s'=\pm}}^{\infty} \frac{f_{M,s'} - f_{N,s}}{\mathcal{E}_{N,s} - \mathcal{E}_{M,s'}} \times \frac{\Omega + \mathcal{E}_{M,s'} - \mathcal{E}_{N,s}}{(\Omega + \mathcal{E}_{M,s'} - \mathcal{E}_{N,s})^2 + \Gamma^2} \times [\mathcal{F}(Ns; Ms')\delta_{N,M-1} - \mathcal{F}(Ms'; Ns)\delta_{M,N-1}], \quad (30)$$

and

$$\text{Im} \left\{ \frac{\sigma_{xy}(\Omega)}{e^2/\hbar} \right\} = \frac{E_1^2}{2\pi} \sum_{\substack{N,M=0 \\ s,s'=\pm}}^{\infty} \frac{f_{M,s'} - f_{N,s}}{\mathcal{E}_{N,s} - \mathcal{E}_{M,s'}} \times \frac{-\Gamma}{(\Omega + \mathcal{E}_{M,s'} - \mathcal{E}_{N,s})^2 + \Gamma^2} \times [\mathcal{F}(Ns; Ms')\delta_{N,M-1} - \mathcal{F}(Ms'; Ns)\delta_{M,N-1}]. \quad (31)$$

For the gapped graphene results, the reader is referred to Eqns. (12)-(15) in Ref.<sup>38</sup>. Note that an overall minus sign should be included in Eqns. (14) and (15) of Ref.<sup>38</sup>. To obtain the appropriate limit,  $\Delta_{\text{so}} = 0$  and  $\Delta_z/2 = \Delta$ . As previously mentioned, a momentum cutoff must be applied to the band structure to ensure that the valence band does not bend back across the zero energy axis. When a magnetic field is applied, a corresponding cutoff on  $N$  must be applied to ensure that  $\mathcal{E}_{N,-} < 0$ .

We begin our discussion by consider the effect of finite  $\Delta$  on the absorptive part of the longitudinal response. This is shown in Fig. 5 for a TI.  $\text{Re}\sigma_{xx}(\Omega)$  is plotted for  $\mu = 0$ ,  $g = 0$  and three values of  $\Delta$ : 0 (solid black), 0.5 meV (dashed blue) and 1 meV (dash-dotted red). The inset shows a schematic plot of the lowest LLs for the different  $\Delta$ 's. The lowest optical transition is marked by the arrows. For  $\Delta = 0$ , the  $N = 0$  LL sits at  $E_0/2$  (for  $g = 0$ ). Therefore, the lowest transition occurs between the  $N = 1$ ,  $s = -$  and  $N=0$  levels. For finite  $\Delta$  less than  $E_0/2$ , the  $N = 0$  level is still at positive energy but its magnitude has decreased; this causes the lowest absorption peak to move down in  $\Omega$ . As  $\Delta$  becomes larger than  $E_0/2$ , the  $N = 0$  level moves to negative energy and the first transition is now between the zeroth level and the  $N = 1$  LL of the conduction band. Now, the frequency of the first transition continues to increase with  $\Delta$  as  $\mathcal{E}_{0,+}$  is pushed further down in energy. Higher optical transitions are present and occur in pairs. This is a signature of the particle-hole asymmetry as the energy for the  $\mathcal{E}_{N,-}$  to  $\mathcal{E}_{N\pm 1,+}$  transition is not the same as the  $\mathcal{E}_{N\pm 1,-}$  to  $\mathcal{E}_{N,+}$  transition. For graphene, the two sets of split higher-energy peaks in Fig. 5 would each coalesce into a single line. It is the presence of the Schrödinger magnetic energy  $E_0$  which sets the scale for the splitting. This energy increases linearly with magnetic field and inversely with decreasing  $m$  as discussed by Li and Carbotte<sup>17</sup>.

Next, we explore the results for finite chemical potential and  $g = 0$ . In Fig. 6, the effect of positive and neg-

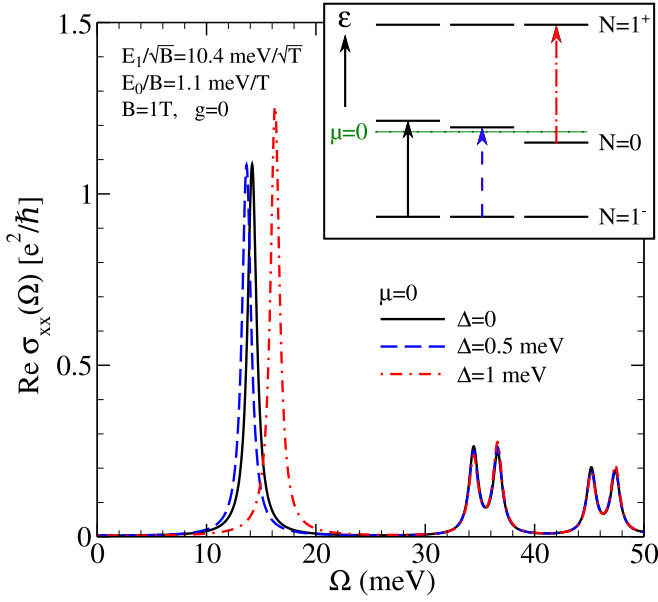


FIG. 5. (Color online)  $\mu = 0$  longitudinal conductivity in a TI for varying  $\Delta$ . The energy of the first absorption process is schematically illustrated by the arrows in the inset. As  $\Delta$  increases, the zeroth LL decreases in energy. This shift is evident in the onset frequency of the first optical transition.

ative  $\mu$  is examined. In frame (a), the results of gapless graphene at the  $K$ -point are shown; due to the particle-hole symmetry, the response is identical for  $\pm\mu$ . The optical transitions which lead to this set of absorption lines are shown in Fig. 7(a) where the arrows are color-coded to correspond to Fig. 6(a). Note that the positions of the  $\mathcal{E}_{N,-}$  LLs are at the negative of the  $\mathcal{E}_{N,+}$  levels. For positive  $\mu$ , the five arrows on the left (black) apply; while, for negative  $\mu$ , it is the five green arrows on the right which are relevant. In both cases, each transition has a one-to-one correspondence with the other set. Figure 6(b), displays the results for a TI when  $\Delta = 0$ . Here the finite Schrödinger term breaks particle-hole symmetry. This is clear from the transitions shown in Fig. 7(b). In this case, even for  $\Delta = 0$ , the negative energy levels do not mirror the positive energy set. The  $N = 0$  level is no longer at zero energy but rather has been pushed to positive energy  $E_0/2$ . The  $\mathcal{E}_{N,+}$  energy is also larger than  $|\mathcal{E}_{N,-}|$ . The black arrow between  $N = 0$  and  $N = 1^+$  (which applies to the lowest line for positive  $\mu$ ) is longer by  $E_0/2$  than the arrows between the  $N = 1^-$  and  $N = 0$  levels which is the first transition when  $\mu$  is negative. Thus, in Fig. 6(b), the first peak of the negative  $\mu$  response (purple) is lower in  $\Omega$  than the corresponding positive  $\mu$  peak (black). The results for a gapped TI are shown in frame (c) of Fig. 6 with the corresponding optical transitions shown in Fig. 7(c). Again, an obvious asymmetry is present between the  $\pm\mu$  regimes. The asymmetry is now much larger than that shown in frame (b) for two reasons: the finite gap adds asymmetry; the larger value

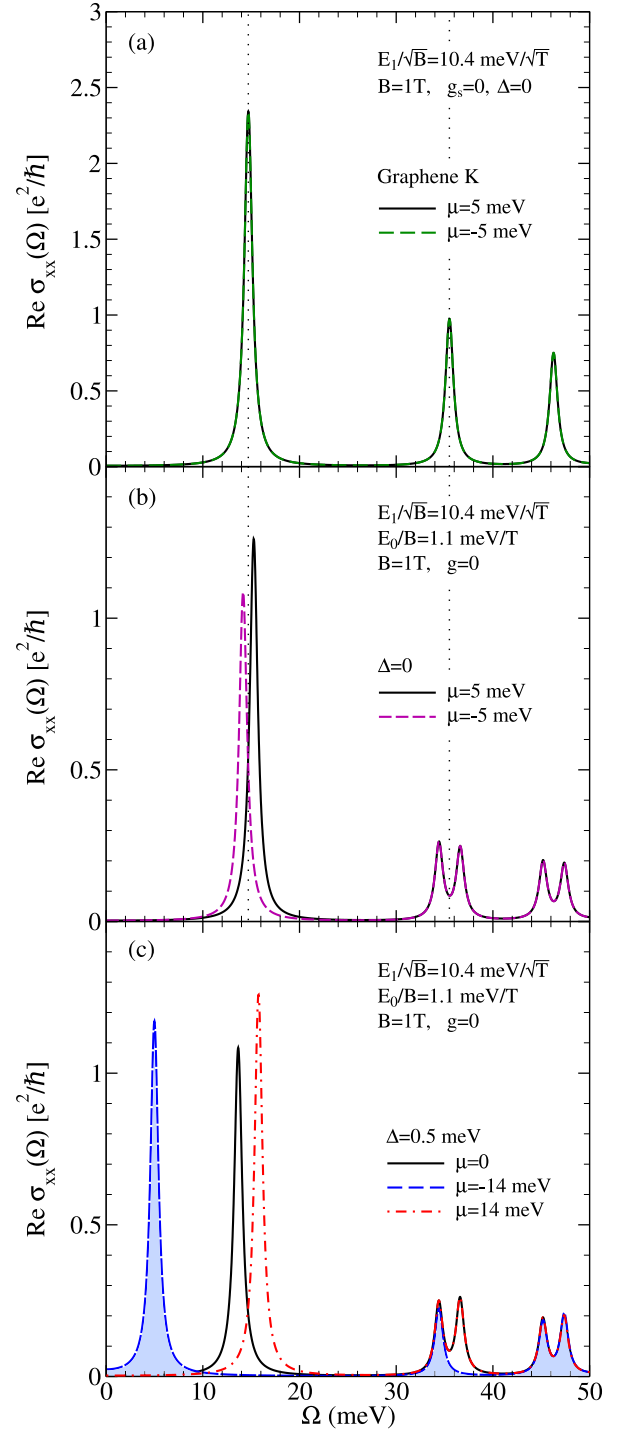


FIG. 6. (Color online) Longitudinal conductivity in (a) gapless graphene, and (b) a gapless and (c) gapped TI for positive and negative  $\mu$ . (a) For graphene, particle-hole symmetry ensures the  $\pm\mu$  results are identical. For a TI [(b)-(c)], the  $\pm\mu$  results are different which emphasizes the asymmetry of the LLs. The shading under the dashed blue curve is for emphasis. Note the missing peak in this case.

of chemical potential enhances it is well. An additional



effect which needs to be emphasized is that one of the absorption lines for  $\mu < 0$  (blue) is missing in the first set of split peaks shown at higher energy. Of the four peaks, the first, third and fourth apply to all three curves ( $\mu = 0$  and  $\mu = \pm 14$  meV), but the second is only present in the black and red curves. The transitions which give rise to these plots are shown in Fig. 7 where the missing transition is indicated by a blue  $\times$ . It cannot occur because the  $N = 1^-$  level is unoccupied.

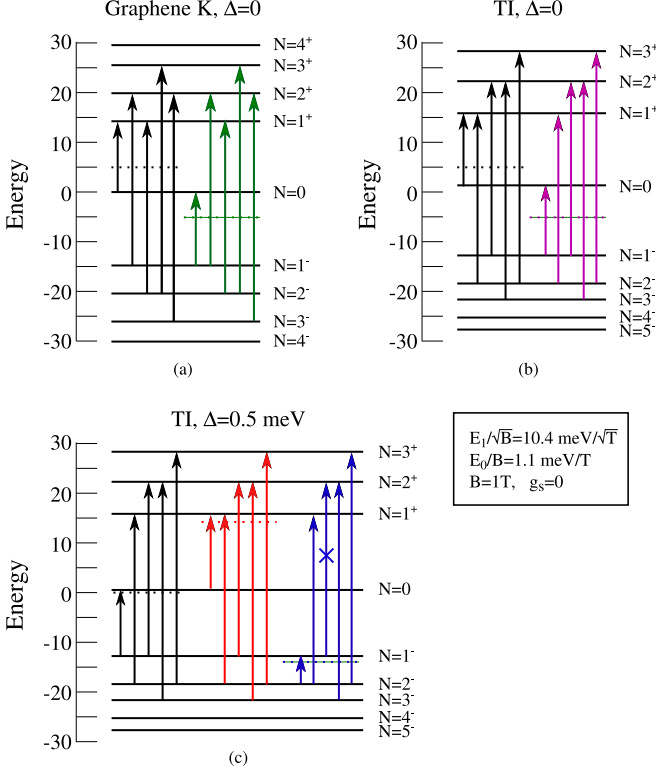


FIG. 7. (Color online) Absorption processes which give rise to the conductivity curves of Fig. 6. The arrows are coloured to correspond with the curves in Fig. 6 and the location of  $\mu$  is given by the dotted lines. For  $\mu = -14$  meV in the gapped TI [blue arrows in frame (c)], the transition from  $\mathcal{E}_{1,-}$  to  $\mathcal{E}_{2,+}$  is forbidden [note the missing absorption line in Fig. 6(c)].

The absorptive part of the transverse Hall conductivity is also of interest and is shown in Fig. 8 for the same parameters as Fig. 6(c). When the optical transition is from the  $N^{\text{th}}$  level in the valence band to the  $(N-1)^{\text{th}}$  LL, the response is negative; while the transition from the  $N^{\text{th}}$  to  $(N+1)^{\text{th}}$  level is positive. This will have important ramifications on the circular dichroism. The shading under the dashed blue curve for negative  $\mu$  again helps to emphasize the missing peak around  $\sim 36.6$  meV which occurs for  $\mu = 0$  (black) and  $\mu = 14$  meV (red). The other three peaks exist for all  $\mu$  considered here. Note that, expect for the lowest energy peak in the dashed curve (blue) and in the dash-dotted curve (red), all other peaks would not exist in the pure Dirac limit. Their existence depends on the presence of the subdominant

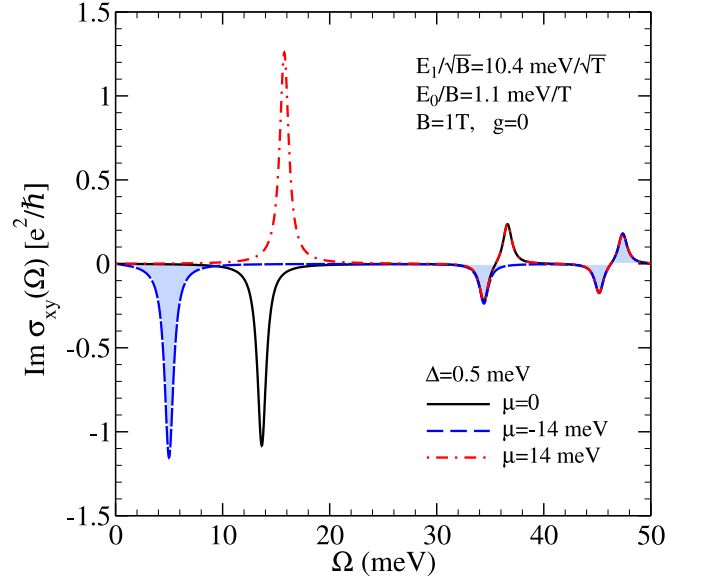


FIG. 8. (Color online) The absorptive part of the transverse Hall conductivity for a TI with parameters set to match Fig. 6(c). For transitions from  $N$  to  $N-1$ , the response is negative; for  $N$  to  $N+1$ , the conductivity is positive [see Fig. 7(c)].

Schrödinger contribution in our Hamiltonian, even if it is small. This represents a qualitative difference between the physics of the two cases.

In the context of TI thin films, both surfaces become important; the two surfaces are marked by an opposite sign of  $\Delta$ . In Fig. 9, the longitudinal conductivity for  $\Delta = \pm 5$  meV is shown. The combined result that would be measured in an optics experiment is given by the solid black curve. Frame (a) corresponds to charge neutrality ( $\mu = 0$ ) while finite chemical potential is considered in frame (b). Aside from an energy shift in the peak associated with  $N = 0$ , we note that the spectral weight for the  $N \rightarrow N \pm 1$  transitions (doublets of higher energy peaks) is markedly different depending on the sign of the gap. In the dash-dotted blue curve, the first peak of the doublet has the largest spectral weight while it is opposite for the dashed red curve.

### A. Circular Dichroism

The response to circularly polarized light is given by  $\sigma_{xx}(\Omega) \pm i\sigma_{xy}(\Omega)$  for right- and left-handed polarization, respectively. Therefore, the absorptive part is determined by

$$\text{Re}\sigma_{\pm}(\Omega) = \text{Re}\sigma_{xx}(\Omega) \mp \text{Im}\sigma_{xy}(\Omega). \quad (32)$$

This is readily evaluated by utilizing Eqns. (27) and (31). A plot of the circular dichroism is given in Fig. 10 for a gapped TI. Frames (a) and (b) show the response



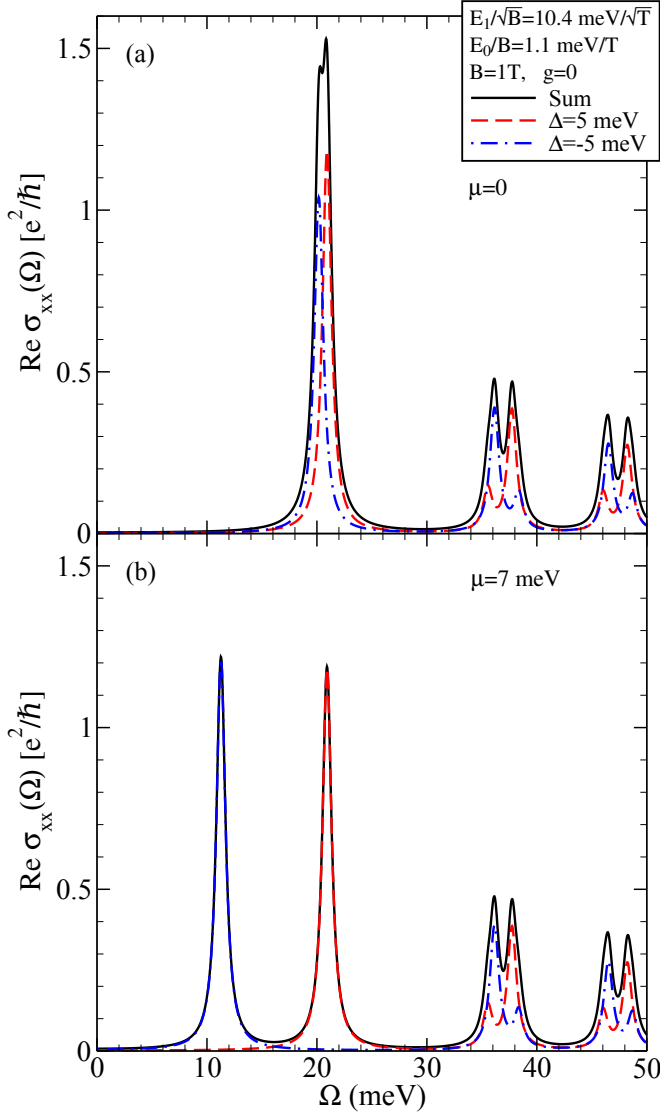


FIG. 9. (Color online) The absorptive part of the longitudinal conductivity for a TI with  $\Delta = \pm 5$  meV for (a)  $\mu = 0$  and (b)  $\mu = 7$  meV.

to right- and left-handed polarized light, respectively. Right-handed light selects out the transitions between the  $N^{\text{th}}$  level in the valence band and the  $(N - 1)^{\text{th}}$  LL in the conduction band. Left-handed light selects the  $N$  to  $N + 1$  transitions. Right circularly polarized light shows a low-energy absorption peak at  $\mu = 0$  (solid black) which moves to lower energy in the dashed blue curve for  $\mu = -14$  meV [frame (a)]. Such a peak is missing for  $\mu = +14$  meV (dash-dotted red). Instead, this peak is present for left circular polarization [frame (b)] while the other two values of  $\mu$  have no such absorption feature. The pairs of peaks shown at higher energy in both frames are present for all three values of  $\mu$  for right-handed light. For left-handed polarization, the higher peak in the pair is present for all  $\mu$  considered here, the lower one is miss-

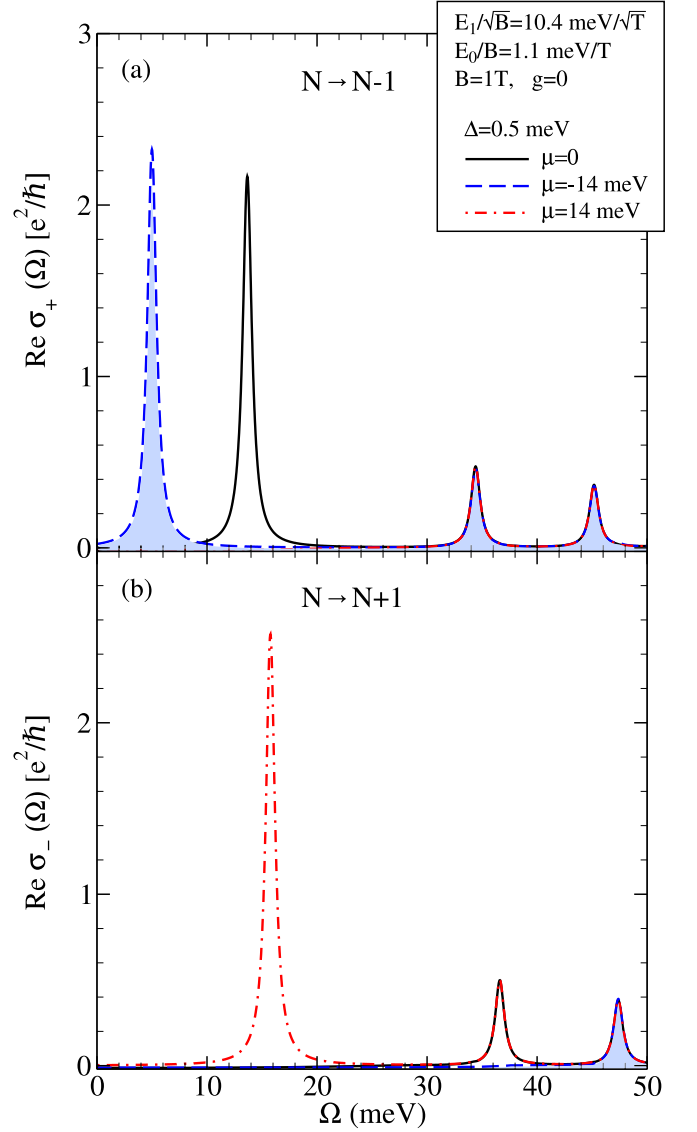


FIG. 10. (Color online) Conductivity response to (a) right- and (b) left-handed circularly-polarized light. (a) For right-handed light, only transitions from  $N$  to  $N - 1$  are present. (b) In the response to left-handed polarization, only the  $N$  to  $N + 1$  processes occur.

ing for  $\mu = 14$  meV (dashed blue). This can be traced to the forbidden transition shown in Fig. 7(c) (marked by the blue  $\times$ ).

## V. MAGNETIZATION OF THE METALLIC SURFACE STATES

To discuss the magnetization of the surface states, we turn to the grand thermodynamic potential

$$\Omega(T, \mu) = -T \int_{-\infty}^{\infty} N(\omega) \ln \left( 1 + e^{(\mu - \omega)/T} \right) d\omega, \quad (33)$$

where  $T$  is the temperature (in units of  $k_B$ ), and  $N(\omega)$  is given by Eqn. (21). The quantities discussed herein, depend on a  $B$  derivative of the grand potential; for convenience, we add  $(\mu/2) \int_{-\infty}^{\infty} N(\omega) d\omega$  to Eqn. (33). The integral of the density of states over all energies gives the total number of states (which must be independent of  $B$ ). Thus, this term will not contribute to the magnetization  $(-\partial\Omega/\partial B)$ . At zero temperature, Eqn. (33) becomes

$$\Omega(\mu) = \int_{-\infty}^0 \left( \omega - \frac{\mu}{2} \right) N(\omega) d\omega + \int_0^{\mu} (\omega - \mu) N(\omega) d\omega + \frac{\mu}{2} \int_0^{\infty} N(\omega) d\omega. \quad (34)$$

For a gapped TI,

$$\int_0^{\infty} N(\omega) d\omega = \int_{-\infty}^0 N(\omega) d\omega + \frac{eB}{h} \Upsilon, \quad (35)$$

where

$$\Upsilon = \begin{cases} -1, & (1+g)E_0/2 < \Delta \\ 0, & (1+g)E_0/2 = \Delta \\ 1, & (1+g)E_0/2 > \Delta \end{cases}. \quad (36)$$

Therefore,

$$\Omega(\mu) = \int_0^{\mu} (\omega - \mu) N(\omega) d\omega + \frac{eB\mu}{2h} \Upsilon + \int_{-\infty}^0 \omega N(\omega) d\omega. \quad (37)$$

The final term does not depend on  $\mu$  and gives the vacuum contribution. This will simply provide a constant background to the  $\mu$  dependence of the magnetization. Keeping only the  $\mu$  dependent pieces, the first two terms of Eqn. (37) give

$$\tilde{\Omega}(\mu) = \frac{eB}{h} \left[ \text{sgn}(\mathcal{E}_{0,+}) (\mathcal{E}_{0,+} - \mu) \Theta(|\mu| - |\mathcal{E}_{0,+}|) \Theta(\text{sgn}(\mathcal{E}_{0,+})\mu) + \frac{\Upsilon\mu}{2} + \sum_{N=1}^{\infty} (\mathcal{E}_{N,+} - \mu) \Theta(\mu - \mathcal{E}_{N,+}) - \sum_{N=1}^{\infty} (\mathcal{E}_{N,-} - \mu) \Theta(\mathcal{E}_{N,-} - \mu) \right], \quad (38)$$

where again, we require all the  $s = -$  states to be negative. Note that  $\Theta(0) \equiv 1/2$  as only half the  $\delta$ -function situated at  $\omega = 0$  is integrated.

The slope of the magnetization is of particular interest as it is related to the quantized Hall conductivity through the Streda relation<sup>20</sup> ( $\sigma_H = e\partial M/\partial\mu$ ). To see the quantization of the slope, note that

$$\frac{\partial\Omega}{\partial\mu} = -\frac{eB}{h} \left[ \text{sgn}(\mathcal{E}_{0,+}) \Theta(|\mu| - |\mathcal{E}_{0,+}|) \Theta(\text{sgn}(\mathcal{E}_{0,+})\mu) - \frac{\Upsilon}{2} + \sum_{N=1}^{\infty} \Theta(\mu - \mathcal{E}_{N,+}) - \Theta(\mathcal{E}_{N,-} - \mu) \right]. \quad (39)$$

The slope of the magnetization is then

$$\frac{\partial M}{\partial\mu} = -\frac{\partial}{\partial B} \frac{\partial\Omega}{\partial\mu} = \frac{e}{h} \left[ \text{sgn}(\mathcal{E}_{0,+}) \Theta(|\mu| - |\mathcal{E}_{0,+}|) \Theta(\text{sgn}(\mathcal{E}_{0,+})\mu) - \frac{\Upsilon}{2} + \sum_{N=1}^{\infty} \Theta(\mu - \mathcal{E}_{N,+}) - \Theta(\mathcal{E}_{N,-} - \mu) \right]. \quad (40)$$

A plot of the magnetization calculated from Eqn. (38) through  $M = -\partial\Omega/\partial B$  is shown in Fig. 11(a). The corresponding slope [see Eqn. (40)] is given in Fig. 11(b). Here,  $g$  is taken to be 1 and the effect of varying  $\Delta$  is emphasized. A saw-tooth oscillation pattern is present in  $M(\mu)$  with the location of the teeth sitting at the various LL energies. The derivative of  $M(\mu)$  gives the quantization of the Hall conductivity. A half-integer quantization is present which is characteristic of Dirac systems. The Hall conductivity is  $\sigma_H = (e^2/h)\nu$  with filling factors  $\nu = \pm 1/2, \pm 3/2, \pm 5/2, \dots$ . For zero gap (solid black curve), the Hall conductivity at  $\mu = 0$  has  $\nu = -1/2$ . This results from the finite value of the  $N = 0$  level<sup>50</sup>. As  $\Delta$  is increased, the location of the  $N = 0$  step moves lower in  $\mu$ . At  $\mu = (1+g)E_0/2 - \Delta$ , the step occurs at zero chemical potential; for larger  $\Delta$ , the small  $|\mu|$  value of  $\sigma_H$  is given by  $\nu = 1/2$ . Again, asymmetry is seen between the negative and positive  $\mu$  regimes. These results have been verified by taking the DC limit ( $\Omega \rightarrow 0$ ) of  $\text{Re}\sigma_{xy}(\Omega)$  [see Eqn. (30)].

To compare this with gapped graphene, the reader is referred to Ref.<sup>39</sup>. The quantization of the Hall conductivity is given by their Eqn. (15). To obtain the gapped graphene result, one must take  $\Delta_{\text{so}} = 0$  and  $\Delta_z/2 \rightarrow \Delta$ . In Fig. 12, the slope of the magnetization is shown for graphene with  $\Delta = 2$  meV,  $E_1/\sqrt{B} = 25.64$  meV/ $\sqrt{\text{T}}$  (a characteristic value for graphene) and  $B = 1\text{T}$ . The upper two frames show the contribution from the  $K$ - and  $K'$ -point, respectively. The lower frame shows the total result. Note that the spin-degeneracy is included. For a single spin species, the Hall response is characterized by a half integer filling factor. The location of the  $N = 0$  LL is symmetric between the two valleys (sitting at  $-\Delta$  for  $K$  and  $\Delta$  for  $K'$ ). As a result, the total system near  $\mu = 0$  is insulating ( $\sigma_H = 0$ ) since the individual valleys contribute equal and opposite edge channels. Here, a Zeeman interaction (dashed purple curve in the upper frame) splits the steps in the conductivity into two.

## VI. MAGNETIC OSCILLATIONS

We now wish to explore the quantum oscillations which exist for low  $B$  fields. This can be done by extracting the oscillating part of the density of states, using that to find the grand potential and taking the appropriate derivative ( $M = -\partial\Omega/\partial B$ ). To leading order in  $1/(mv_F^2)$ , the

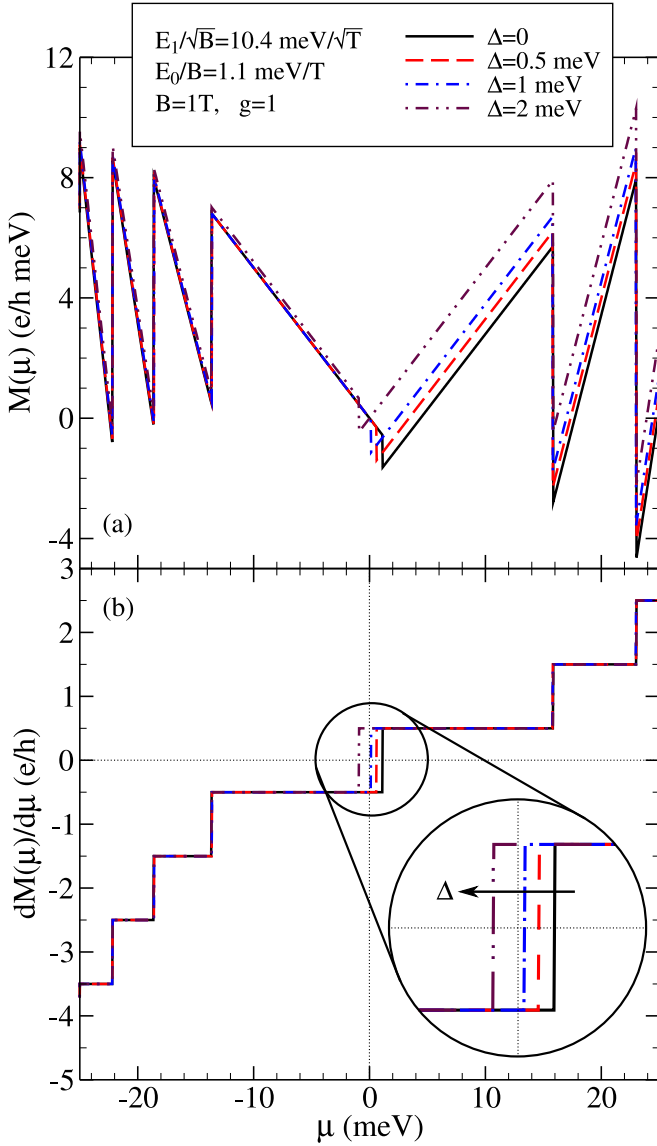


FIG. 11. (Color online) (a)  $\mu$  dependence of the magnetization for a TI with varying  $\Delta$ . (b) The corresponding slope of the magnetization (which is related to the Hall conductivity through the Streda relation:  $\partial M/\partial \mu = (1/e)\sigma_H$ ). In all cases, a half-integer quantization is present; however, the step to  $+1/2$  can be tuned from positive to negative  $\mu$  by increasing  $\Delta$ .

magnetization is (see Appendix)

$$M_{\text{osc}}^k(\mu) \approx -\frac{e}{2\pi k h} \frac{(\mu^2 - \Delta^2)}{\mu} \left[ 1 + \frac{\mu^2 - \Delta^2}{2\mu m v_F^2} \right] \sin(2\pi k x_1), \quad (41)$$

where

$$x_1 \approx \frac{\left(1 - \frac{\mu}{m v_F^2}\right) [\mu^2 - \Delta^2]}{2E_1^2} + \frac{\Delta(1+g)}{2m v_F^2}. \quad (42)$$

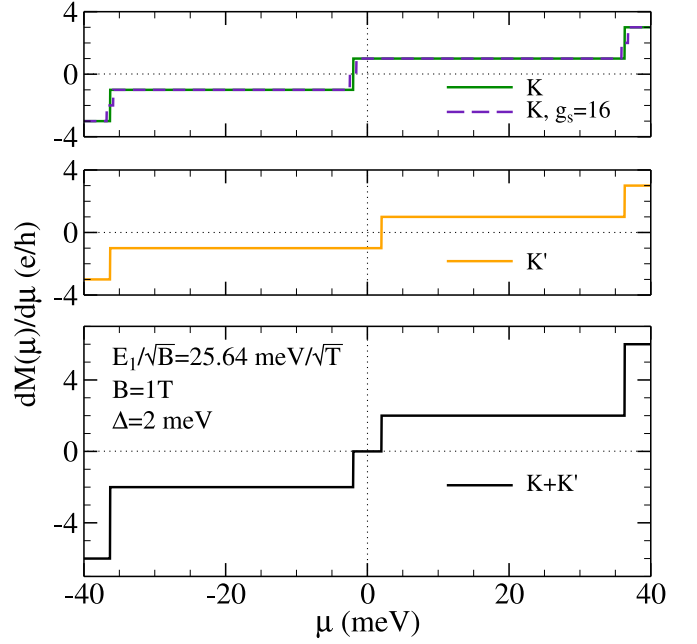


FIG. 12. (Color online) The slope of the magnetization for gapped graphene at (top) the  $K$ -point, (middle) the  $K'$ -point and (lower) the total response. The step which results from the  $N = 0$  level is at  $\mu = \pm\Delta$  for the  $K$ - and  $K'$ -point, respectively. Thus, the total system is insulating for  $-\Delta < \mu < \Delta$ . Zeeman (dashed purple curve in the top frame) splits the steps into two.

Comparing Eqn. (42) to the customary<sup>23</sup>

$$x_1 = \frac{\hbar A(\mu)}{2\pi e B} - \gamma, \quad (43)$$

the coefficient of the  $E_1^{-2} \propto 1/B$  term in Eqn. (42) is in fact related to the area of the cyclotron orbit:

$$A(\mu) = \pi k_F^2 \approx \frac{\pi [\mu^2 - \Delta^2]}{\hbar^2 v_F^2} \left( 1 - \frac{\mu}{m v_F^2} \right). \quad (44)$$

The remainder is the phase shift which is independent of  $B$ . It is,

$$\gamma = -\frac{\Delta(1+g)}{2m v_F^2}. \quad (45)$$

In the pure Dirac limit ( $m \rightarrow \infty$ ,  $\Delta \rightarrow 0$ ),  $A(\mu)$  reduces to the correct value<sup>23,51</sup> of  $\pi\mu^2/(\hbar^2 v_F^2)$ . For finite  $m$  and  $\Delta = 0$ , we obtain the correction of  $-\pi\mu^2/(\hbar^2 v_F^2)[\mu/(m v_F^2)]$  to  $A(\mu)$ <sup>50</sup>. Clearly, the gapless limit has a phase shift of 0 associated with a Berry's phase of  $\pi$ <sup>44,50</sup>. The amplitude of the quantum oscillations [in units of  $-e/(2\pi k h)$ ] is

$$\mathcal{A}_M = \frac{(\mu^2 - \Delta^2)}{\mu} \left[ 1 + \frac{\mu^2 - \Delta^2}{2\mu m v_F^2} \right] \quad (46)$$

which properly reduces to the results of Sharapov et al<sup>52</sup> when  $m \rightarrow \infty$ , i.e.  $\mathcal{A}_M = (\mu^2 - \Delta^2)/\mu$  (see their Eqn. (8.10) in the pure limit), and to the results of Tabert and Carbotte<sup>50</sup> when  $\Delta = 0$ , i.e.  $\mathcal{A}_M = \mu[1 + \mu/(2mv_F^2)]$  [see their Eqn. (49)]. Wright and McKenzie<sup>22</sup> applied the semiclassical quantization method of Onsager<sup>30</sup>. This was augmented by including a first correction in the band structure for the energy shift due to the magnetic response of the system (as discussed by Fuchs et al<sup>24</sup>). They obtained a phase offset in the magnetic oscillations which, to within a sign that can be traced to the chirality of the Hamiltonian, reduces to our result [Eqn. (45)] for large  $m$  and when the Zeeman interaction is neglected (i.e.  $g = 0$ ).

## VII. CONCLUSIONS

The simplest description of the surface states of a TI is a single Dirac cone centred at the  $\Gamma$ -point of the 2D surface Brillouin zone. The presence of a subdominant quadratic-in-momentum Schrödinger term to the purely relativistic linear-in-momentum Dirac Hamiltonian is also an important feature; it reshapes the perfect graphene-like cone into an hourglass shape with the narrowing of the conduction cone and outward fanning of the lower cone (valence band). Particle-hole symmetry is lost which can have important ramifications on physical properties such as the density of states, magnetization and the optical properties (both AC and DC).

Doping the surface of a TI with magnetic atoms breaks time-reversal symmetry and creates a gap of  $2\Delta$  at the Dirac point. Alternatively, a slab can be made thin enough that the top and bottom surface states hybridize; consequently, both become gapped. The sign of the gap is opposite on the two surfaces. We have studied the magneto-optical response as a function of energy of such systems including as well the effect of a Zeeman interaction. Particular emphasis is given to the particle-hole asymmetry brought about by the Schrödinger mass term associated with the non-relativistic piece of the Hamiltonian. In a finite magnetic field, both the Zeeman interaction and the gap can modify this asymmetry. Comparing the integrated density of states  $I(\omega_{\max})$  to  $\omega_{\max}$  in a TI to a single valley of graphene highlights the important differences between the two cases. Namely, particle-hole asymmetry and the very different effect of Zeeman splitting. For a TI, Zeeman coupling does not split the steps of  $I(\omega_{\max})$  into two spin-polarized substeps displaced by a constant amount as in graphene. Instead, it shifts the onset of the various steps by a level-dependent amount. For a single valley of graphene, the gap does produce asymmetry; but, this will not be seen if the two valleys of opposite chirality are superimposed. Since the Hamiltonian for a TI involves real spin (as opposed to pseudospin in graphene and related materials such as MoS<sub>2</sub> and silicene), only the  $N = 0$  level is fully spin-polarized<sup>16</sup>. All other LLs are found to have a much reduced value of the

average  $\hat{s}_z \sim (\hbar/2)[\Delta - (1+g)E_0/2]/(E_1\sqrt{2N})$ . This is to be contrasted with the pseudospin case where all levels are fully  $\hat{s}_z$  polarized with  $s_z = \pm\hbar/2$  depending on whether the LL has moved up or down in energy due to Zeeman.

The dynamic longitudinal magneto-optical conductivity for a TI also displays asymmetry between positive and negative values of the chemical potential  $\mu$ . Even for  $\mu = 0$ , the absorption line associated with the  $\mathcal{E}_{1,-}$  to  $\mathcal{E}_{0,+}$  transition can be eliminated and replaced by a  $\mathcal{E}_{0,+}$  to  $\mathcal{E}_{1,+}$  line by manipulating the gap value. There are peaks in the imaginary part of the transverse Hall AC conductivity which do not exist for the pure Dirac case. These features do not depend on the subdominant non-relativistic term being large. They correspond to the interband optical transitions and come as pairs of peaks (one positive the other negative). They translate into new absorption peaks for circularly polarized light.

While the starting formulas for the magnetization and the DC limit for the real part of the Hall conductivity are quite different, the values of the Hall plateaus that ultimately result are identical and agree with those for the pure relativistic case. Although, the values of chemical potential at which a new step appears are different. They depend on the presence of a non-relativistic piece in the Hamiltonian. For example, the transition from  $-1/2$  to  $1/2$  no longer corresponds to zero chemical potential. This transition is shifted to the position of the  $N = 0$  LL which is given by  $(1+g)E_0/2 - \Delta$ . We emphasize that this quantity remains finite when both the gap ( $\Delta$ ) and Zeeman term ( $g$ ) are zero. The shift of  $E_0/2$  is directly due to the non-relativistic Schrödinger term in our Hamiltonian. Even if this term is small, it provides a qualitative modification of the physics as compared with the pure relativistic case. Note that including a Zeeman term and the gap further shifts the energy at which the Hall conductivity transitions from  $-1/2$  to  $1/2$ . It also depends on the sign of the gap. For  $\Delta > 0$ , the shift is to lower energies while for  $\Delta < 0$ , it is toward higher energies.

We consider the quantum oscillations that arise in the magnetization. Unlike previous work, our considerations do not involve any semiclassical arguments based on Onsager's quantization condition for cyclotron orbits. Here we proceed from the grand potential and use a Poisson formula to obtain the low-field limit ( $B \rightarrow 0$ ). We find that when the gap is finite and a Schrödinger mass term is also included, there is an offset in the phase shift associated with the magnetic oscillations. To lowest order, it is given by  $\gamma \equiv -\Delta(1+g)/(2mv_F^2)$  which reduces to zero when  $\Delta = 0$  or  $m \rightarrow \infty$ . In both these cases,  $\gamma = 0$  as expected for Dirac fermions. We also see a dependence on the Zeeman splitting ( $g$ ). In this formula,  $v_F$  is the Dirac Fermi velocity. Except for a sign due to the chirality of the Hamiltonian, this agrees with Eqn. (26) in Ref.<sup>22</sup> when only the leading order in  $1/m$  is retained. These authors employed a semiclassical approximation to obtain their results. We note that this phase offset remains

even though the quantization of the Hall plateaus is relativistic ( $\pm 1/2, \pm 3/2, \pm 5/2, \dots$ ). A new expression for the amplitude of the magnetic oscillations is given which properly reduces to that of gapped graphene when the subdominant Schrödinger term is dropped and to that previously found in Ref.<sup>50</sup> when the Schrödinger mass is included but the gap is set to zero.

## ACKNOWLEDGMENTS

We thank E. J. Nicol for support of this work. This work has been supported by the Natural Sciences and Engineering Research Council of Canada and, in part, by the Canadian Institute for Advanced Research.

## Appendix

To obtain the magnetic oscillations, return to Eqn. (21) and express it as

$$N(\omega) = \frac{eB}{h} \frac{d}{d\omega} \left[ \Theta(\omega - \mathcal{E}_{0,+}) + \sum_{\substack{N=1 \\ s=\pm}}^{\infty} \Theta(\omega - \mathcal{E}_{N,s}) \right]. \quad (\text{A.1})$$

The oscillating part of the density of states can be extracted by applying the Poisson formula

$$\begin{aligned} \sum_{N=1}^{\infty} F(N) &= -\frac{1}{2}F(0) + \int_0^{\infty} F(x)dx \\ &+ 2 \sum_{k=1}^{\infty} \int_0^{\infty} F(x) \cos(2\pi kx) dx. \end{aligned} \quad (\text{A.2})$$

Following arguments given by Suprunenko et al<sup>51</sup>, this gives,

$$\begin{aligned} N(\omega) &= \frac{eB}{h} \frac{d}{d\omega} \left\{ \frac{1}{2} \Theta(\omega - |\mathcal{E}_{0,+}|) - \frac{1}{2} \Theta(\omega + |\mathcal{E}_{0,+}|) \right. \\ &+ [\Theta(\omega + |\mathcal{E}_{0,+}|) + \Theta(\omega - |\mathcal{E}_{0,+}|) - \Theta(\omega - \omega_{\min})] \\ &\times \left[ x_1 + \sum_{k=1}^{\infty} \frac{1}{\pi k} \sin(2\pi kx_1) \right] \\ &\left. + \Theta(\omega - \omega_{\min}) \left[ x_2 + \sum_{k=1}^{\infty} \frac{1}{\pi k} \sin(2\pi kx_2) \right] \right\}, \end{aligned} \quad (\text{A.3})$$

where

$$\omega_{\min} = -\frac{E_1^2}{2E_0} - \frac{E_0}{2E_1^2} \left( \frac{E_0}{2}(1+g) - \Delta \right)^2, \quad (\text{A.4})$$

and

$$x_i = \frac{E_1^2}{E_0^2} + \frac{\omega}{E_0} + (-)^i \sqrt{\frac{E_1^4}{E_0^4} + \frac{2E_1^2\omega}{E_0^3} + \left( \frac{1+g}{2} - \frac{\Delta}{E_0} \right)^2}, \quad (\text{A.5})$$

with  $i = 1, 2$ . This generalizes Eqn. (8) of Ref.<sup>51</sup> to include a gap and Zeeman splitting. For  $E_1^2 \gg E_0^2$  and  $m \rightarrow \infty$ ,

$$x_1 \approx \frac{\omega^2}{2E_1^2} \left[ 1 - \frac{\omega}{mv_F^2} \right] - \frac{bE_0}{2mv_F^2} + \frac{\omega bE_0}{(mv_F^2)^2}, \quad (\text{A.6})$$

where

$$b \equiv \frac{[\Delta - (1+g)E_0/2]^2}{E_0^2}, \quad (\text{A.7})$$

and, thus

$$\begin{aligned} -\frac{bE_0}{2mv_F^2} &= -\frac{1}{2mv_F^2 E_0} \left[ \Delta^2 - E_0 \Delta (1+g) \right. \\ &\left. + \left( \frac{E_0}{2} \right)^2 (1+g)^2 \right]. \end{aligned} \quad (\text{A.8})$$

The last term in the above expression goes like  $E_0 \propto B$  and will drop out in the limit  $B \rightarrow 0$ <sup>50</sup>. This leaves

$$-\frac{bE_0}{2mv_F^2} = -\frac{\Delta^2}{2E_1^2} + \frac{\Delta(1+g)}{2mv_F^2}, \quad (\text{A.9})$$

where the last term is constant in  $B$  and will therefore contribute a constant phase to the quantum oscillations. We now have

$$x_1 \approx \frac{1}{2E_1^2} \left[ 1 - \frac{\omega}{mv_F^2} \right] [\omega^2 - \Delta^2] + \frac{\Delta(1+g)}{2mv_F^2} \left[ 1 - \frac{\omega}{mv_F^2} \right]. \quad (\text{A.10})$$

To lowest order in  $1/(mv_F^2)$ ,

$$x_1 \approx \frac{1}{2E_1^2} \left[ 1 - \frac{\omega}{mv_F^2} \right] [\omega^2 - \Delta^2] + \frac{\Delta(1+g)}{2mv_F^2}. \quad (\text{A.11})$$

As we do not allow the valence band to bend back toward the zero energy axis, and we focus on  $\omega > 0$ ,  $x_2$  can be ignored and the oscillating part of the density of states is entirely determined by

$$N_{\text{osc}}(\omega) = \frac{eB}{h} \frac{d}{d\omega} \Theta(\omega - |\mathcal{E}_{0,+}|) \sum_{k=1}^{\infty} \frac{1}{\pi k} \sin(2\pi kx_1). \quad (\text{A.12})$$

All the information about the quantum oscillations is contained in the first term of Eqn. (37). The oscillating part of the grand thermodynamic potential is thus,

$$\Omega_{\text{osc}}(\mu) = \int_0^{\mu} (\omega - \mu) N_{\text{osc}}(\omega) d\omega. \quad (\text{A.13})$$

where

$$N_{\text{osc}}(\omega) = \frac{eB}{h} \sum_{k=1}^{\infty} \frac{d}{d\omega} N_{\text{osc}}^k(\omega), \quad (\text{A.14})$$

and

$$N_{\text{osc}}^k(\omega) = \Theta(\omega - |\mathcal{E}_{0,+}|) \frac{\sin(2\pi k x_1)}{\pi k}. \quad (\text{A.15})$$

Therefore,

$$\Omega_{\text{osc}}^k(\mu) = -\frac{eB}{\pi k h} \int_0^\mu \Theta(\omega - |E_{0,+}|) \sin(2\pi k x_1) d\omega. \quad (\text{A.16})$$

The magnetization is given by  $M = -\partial\Omega/\partial B$ . Using

$$\frac{\partial x_1}{\partial B} = -\frac{1}{2E_1^2 B} \left[ 1 - \frac{\omega}{mv_F^2} \right] [\omega^2 - \Delta^2], \quad (\text{A.17})$$

the magnetization can be written as

$$M_{\text{osc}}^k(\mu) = \frac{e}{\pi k h} \int_{|\Delta - (1+g)E_0/2|}^\mu \left[ \sin(2\pi k x_1) - \frac{\pi k}{E_1^2} \left[ 1 - \frac{\omega}{mv_F^2} \right] [\omega^2 - \Delta^2] \cos(2\pi k x_1) \right] d\omega. \quad (\text{A.18})$$

In the limit of interest ( $E_0 \rightarrow 0$ ), the lower bound of integration can be replaced by  $\Delta$ . To solve the integral, define

$$\begin{aligned} y &\equiv \left[ 1 - \frac{\omega}{mv_F^2} \right] [\omega^2 - \Delta^2] \\ \Rightarrow dy &= \left[ 2\omega - \frac{3\omega^2}{mv_F^2} + \frac{\Delta^2}{mv_F^2} \right] d\omega \\ \Rightarrow d\omega &= \frac{dy}{2\omega - \frac{3\omega^2}{mv_F^2} + \frac{\Delta^2}{mv_F^2}}. \end{aligned} \quad (\text{A.19})$$

For  $m \rightarrow \infty$ ,  $y \approx \omega^2 - \Delta^2$  so, to a first order correction in  $1/m$ ,

$$y \approx \left[ 1 - \frac{\sqrt{y + \Delta^2}}{mv_F^2} \right] [\omega^2 - \Delta^2] \quad (\text{A.20})$$

Expanding further, we obtain

$$\omega \approx \sqrt{y \left[ 1 + \frac{\sqrt{y + \Delta^2}}{mv_F^2} \right] + \Delta^2}, \quad (\text{A.21})$$

and

$$\omega^2 \approx y \left[ 1 + \frac{\sqrt{y + \Delta^2}}{mv_F^2} \right] + \Delta^2. \quad (\text{A.22})$$

Therefore,

$$\frac{1}{2\omega - \frac{3\omega^2}{mv_F^2} + \frac{\Delta^2}{mv_F^2}} \approx \frac{1}{2\sqrt{y + \Delta^2}} \left[ 1 + \frac{\sqrt{y + \Delta^2}}{mv_F^2} \right]. \quad (\text{A.23})$$

Consolidating these results,

$$\begin{aligned} M_{\text{osc}}^k(\mu) &\approx \frac{e}{2\pi k h} \int_0^\alpha \left\{ \sin \left( 2\pi k \left[ \frac{y}{2E_1^2} + \delta \right] \right) - \frac{\pi k y}{E_1^2} \cos \left( 2\pi k \left[ \frac{y}{2E_1^2} + \delta \right] \right) \right\} \\ &\quad \times \left[ \frac{1}{\sqrt{y + \Delta^2}} + \frac{1}{mv_F^2} \right] dy. \end{aligned} \quad (\text{A.24})$$

where

$$\alpha \equiv \left[ 1 - \frac{\mu}{mv_F^2} \right] [\mu^2 - \Delta^2], \quad (\text{A.25})$$

and

$$\delta \equiv \frac{\Delta(1+g)}{2mv_F^2}. \quad (\text{A.26})$$

Next, define  $y \equiv \alpha x$  to obtain

$$\begin{aligned} M_{\text{osc}}^k(\mu) &\approx \frac{e}{2\pi k h} \int_0^1 dx \left\{ \frac{\alpha}{\sqrt{\alpha x + \Delta^2}} + \frac{\alpha}{mv_F^2} \right\} \\ &\quad \times \left\{ \sin(\alpha x + \bar{\delta}) - \alpha x \cos(\alpha x + \bar{\delta}) \right\}, \end{aligned} \quad (\text{A.27})$$

where  $a \equiv \pi k \alpha / E_1^2$ ,  $\bar{\delta} \equiv 2\pi k \delta$ , and again we work to lowest order in  $1/(mv_F^2)$ . To proceed, consider

$$\begin{aligned} I_1 &\equiv \int_0^1 \left\{ \sin(\alpha x + \bar{\delta}) - \alpha x \cos(\alpha x + \bar{\delta}) \right\} \frac{dx}{\sqrt{x + \Delta^2/\alpha}} \\ &= \int_0^1 \sin(\alpha x + \bar{\delta}) \frac{dx}{\sqrt{x + \Delta^2/\alpha}} \\ &\quad - \int_0^1 \frac{x}{\sqrt{x + \Delta^2/\alpha}} dx \frac{d}{dx} \sin(\alpha x + \bar{\delta}). \end{aligned} \quad (\text{A.28})$$

Integrating the second term by parts and defining  $\omega \equiv x + \Delta^2/\alpha$  and  $b \equiv \bar{\delta} - a\Delta^2/\alpha$ , we obtain

$$\begin{aligned} I_1 &= \int_{\Delta^2/\alpha}^{1+\Delta^2/\alpha} \sin(a\omega + b) \left[ \frac{3}{2\sqrt{\omega}} + \frac{\Delta^2/\alpha}{2\omega^{3/2}} \right] d\omega \\ &\quad - \frac{\sin(a + \bar{\delta})}{\sqrt{1 + \Delta^2/\alpha}}. \end{aligned} \quad (\text{A.29})$$

To solve the remaining integral, we use the definition of the Fresnel sine and cosine integrals:

$$\mathcal{S}(z) \equiv \int_0^z \sin\left(\frac{1}{2}\pi t^2\right) dt, \quad (\text{A.30})$$

and

$$\mathcal{C}(z) \equiv \int_0^z \cos\left(\frac{1}{2}\pi t^2\right) dt, \quad (\text{A.31})$$

respectively. We make use of the fact that

$$\sin(a\omega + b) = \sin(a\omega)\cos b + \cos(a\omega)\sin b. \quad (\text{A.32})$$

We have

$$\begin{aligned} \int_{\Delta^2/\alpha}^{1+\Delta^2/\alpha} \sin(a\omega + b) \frac{d\omega}{\sqrt{\omega}} \\ = \sqrt{\frac{2\pi}{a}} \left[ \mathcal{S} \left( \sqrt{\frac{2a}{\pi}} \omega \right) \Big|_{\sqrt{\Delta^2/\alpha}}^{\sqrt{1+\Delta^2/\alpha}} \cos b \right. \\ \left. + \mathcal{C} \left( \sqrt{\frac{2a}{\pi}} \omega \right) \Big|_{\sqrt{\Delta^2/\alpha}}^{\sqrt{1+\Delta^2/\alpha}} \sin b \right]. \end{aligned} \quad (\text{A.33})$$

Once evaluated at the limits, we will have Fresnel functions with arguments proportional to  $\sqrt{a}$ . We are interested in the limit  $a \rightarrow \infty$ . We use the asymptotic expansions of the Fresnel integrals for  $a \rightarrow \infty$ :

$$\mathcal{S}(\gamma\sqrt{a}) \approx \frac{1}{2} - \frac{1}{\pi\gamma\sqrt{a}} \cos\left(\frac{1}{2}\pi\gamma^2 a\right), \quad (\text{A.34})$$

and

$$\mathcal{C}(\gamma\sqrt{a}) \approx \frac{1}{2} + \frac{1}{\pi\gamma\sqrt{a}} \sin\left(\frac{1}{2}\pi\gamma^2 a\right). \quad (\text{A.35})$$

Therefore, Eqn. (A.33) is proportional to  $1/a$  and is negligible in the limit of interest. We now return to Eqn. (A.29) and consider

$$\begin{aligned} \int_{\Delta^2/\alpha}^{1+\Delta^2/\alpha} \sin(a\omega + b) \frac{d\omega}{\omega^{3/2}} \\ = -\frac{2}{\sqrt{\omega}} \sin(a\omega + b) \Big|_{\Delta^2/\alpha}^{1+\Delta^2/\alpha} \\ - 2\sqrt{2\pi a} \left[ \mathcal{S} \left( \sqrt{\frac{2a}{\pi}} \omega \right) \sin b - \mathcal{C} \left( \sqrt{\frac{2a}{\pi}} \omega \right) \cos b \right] \Big|_{\sqrt{\Delta^2/\alpha}}^{\sqrt{1+\Delta^2/\alpha}}. \end{aligned} \quad (\text{A.36})$$

Applying the Fresnel expansions for  $a \rightarrow \infty$ , this term is zero. We arrive at the simple result

$$I_1 \approx -\frac{\sin(a + \bar{\delta})}{\sqrt{1 + \Delta^2/\alpha}}. \quad (\text{A.37})$$

Next, we consider

$$\begin{aligned} I_2 &\equiv \int_0^1 \{\sin(ax + \bar{\delta}) - a x \cos(ax + \bar{\delta})\} dx \\ &= \int_0^1 \sin(ax + \bar{\delta}) dx - \int_0^1 x dx \frac{d}{dx} \sin(ax + \bar{\delta}). \end{aligned} \quad (\text{A.38})$$

Integrating the second term by parts, we obtain

$$I_2 = -\frac{2}{a} \cos(ax + \bar{\delta}) \Big|_0^1 - \sin(a + \bar{\delta}). \quad (\text{A.39})$$

In the limit  $a \rightarrow \infty$ ,

$$I_2 \approx -\sin(a + \bar{\delta}). \quad (\text{A.40})$$

Combing Eqns. (A.37) and (A.40), the magnetization becomes

$$M_{\text{osc}}^k(\mu) \approx -\frac{e}{2\pi k h} \left[ \frac{\alpha}{\sqrt{\alpha + \Delta^2}} + \frac{\alpha}{m v_F^2} \right] \sin(a + \bar{\delta}). \quad (\text{A.41})$$

Written in terms of the original variables,

$$M_{\text{osc}}^k(\mu) \approx -\frac{e}{2\pi k h} \frac{(\mu^2 - \Delta^2)}{\mu} \left[ 1 + \frac{\mu^2 - \Delta^2}{2\mu m v_F^2} \right] \sin(2\pi k x_1), \quad (\text{A.42})$$

where

$$x_1 \approx \frac{\left(1 - \frac{\mu}{m v_F^2}\right) [\mu^2 - \Delta^2]}{2E_1^2} + \frac{\Delta(1 + g)}{2m v_F^2}. \quad (\text{A.43})$$

- 
- <sup>1</sup> M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).  
<sup>2</sup> X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).  
<sup>3</sup> D. Hsieh, Y. Xia, D. Qian, L. Wray, J. H. Dil, F. Meier, J. Osterwalder, L. Patthey, J. G. Checkelsky, N. P. Ong, A. V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature **460**, 1101 (2009).  
<sup>4</sup> B. Roy, J. D. Sau, M. Dzero, and V. Galitski, Phys. Rev. B **90**, 155314 (2014).  
<sup>5</sup> L. Fu and C. L. Kane, Phys. Rev. Lett. **100**, 096407 (2008).  
<sup>6</sup> K. Pluciski, W. Gruhn, I. Kityk, W. Imioek, H. Kaddouri, and S. Benet, Optics Communications **204**, 355 (2002).  
<sup>7</sup> Y. L. Chen, J.-H. Chu, J. G. Analytis, Z. K. Liu, K. Igarashi, H.-H. Kuo, X. L. Qi, S. K. Mo, R. G.

- Moore, D. H. Lu, M. Hashimoto, T. Sasagawa, S. C. Zhang, I. R. Fisher, Z. Hussain, and Z. X. Shen, Science **329**, 659 (2010).  
<sup>8</sup> H.-Z. Lu, W.-Y. Shan, W. Yao, Q. Niu, and S.-Q. Shen, Phys. Rev. B **81**, 115407 (2010).  
<sup>9</sup> J. Linder, T. Yokoyama, and A. Sudbø, Phys. Rev. B **80**, 205401 (2009).  
<sup>10</sup> W.-Y. Shan, H.-Z. Lu, and S.-Q. Shen, New Journal of Physics **12**, 043048 (2010).  
<sup>11</sup> R. Yoshimi, A. Tsukazaki, Y. Kozuka, J. Falson, K. S. Takahashi, J. G. Checkelsky, N. Nagaosa, M. Kawasaki, and Y. Tokura, Nature Commun. **6**, 6627 (2015).  
<sup>12</sup> S.-B. Zhang, H.-Z. Lu, and S.-Q. Shen, (2015), arXiv:1502.01792.  
<sup>13</sup> M. Neupane, S.-Y. Xu, R. Sankar, N. Alidoust, G. Bian, C. Liu, I. Belopolski, T.-R. Chang, H.-T. Jeng, H. Lin,



- A. Bansil, F. Chou, and M. Z. Hasan, *Nature Commun.* **5**, 3786 (2014).
- <sup>14</sup> M. Tahir, K. Sabeeh, and U. Schwingenschlöggl, *Scientific Reports* **3**, 1261 (2013).
- <sup>15</sup> V. P. Gusynin, V. A. Miransky, S. G. Sharapov, and I. A. Shovkovy, *Phys. Rev. B* **74**, 195429 (2006).
- <sup>16</sup> M. M. Vazifeh and M. Franz, *Phys. Rev. B* **86**, 045451 (2012).
- <sup>17</sup> Z. Li and J. P. Carbotte, *Phys. Rev. B* **88**, 045414 (2013).
- <sup>18</sup> A. D. LaForge, A. Frenzel, B. C. Pursley, T. Lin, X. Liu, J. Shi, and D. N. Basov, *Phys. Rev. B* **81**, 125120 (2010).
- <sup>19</sup> M. Orlita, B. A. Piot, G. Martinez, N. K. S. Kumar, C. Faugeras, M. Potemski, C. Michel, E. M. Hankiewicz, T. Brauner, C. Drašar, S. Schreyeck, S. Grauer, K. Brunner, C. Gould, C. Brüne, and L. W. Molenkamp, *Phys. Rev. Lett.* **114**, 186401 (2015).
- <sup>20</sup> L. Smrcka and P. Streda, *J. Phys. C: Solid State Phys.* **10**, 2153 (1977).
- <sup>21</sup> Z. Wang, Z.-G. Fu, S.-X. Wang, and P. Zhang, *Phys. Rev. B* **82**, 085429 (2010).
- <sup>22</sup> A. R. Wright and R. H. McKenzie, *Phys. Rev. B* **87**, 085411 (2013).
- <sup>23</sup> I. A. Luk'yanchuk and Y. Kopelevich, *Phys. Rev. Lett.* **93**, 166402 (2004).
- <sup>24</sup> J. Fuchs, F. Piechon, M. Goerbig, and G. Montambaux, *Eur. Phys. J. B* **77**, 351 (2010).
- <sup>25</sup> A. A. Taskin, Z. Ren, S. Sasaki, K. Segawa, and Y. Ando, *Phys. Rev. Lett.* **107**, 016801 (2011).
- <sup>26</sup> G. P. Mikitik and Y. V. Sharlai, *Phys. Rev. B* **85**, 033301 (2012).
- <sup>27</sup> Y. Ando, *J. Phys. Soc. Jpn.* **82**, 102001 (2013).
- <sup>28</sup> A. Raoux, M. Morigi, J.-N. Fuchs, F. Piéchon, and G. Montambaux, *Phys. Rev. Lett.* **112**, 026402 (2014).
- <sup>29</sup> K. Kishigi and Y. Hasegawa, *Phys. Rev. B* **90**, 085427 (2014).
- <sup>30</sup> L. Onsager, *Philosophical Magazine* **43**, 1006 (1952).
- <sup>31</sup> Z. Li and J. P. Carbotte, *Phys. Rev. B* **89**, 165420 (2014).
- <sup>32</sup> Y. A. Bychkov and E. I. Rashba, *JETP Lett.* **39**, 78 (1984).
- <sup>33</sup> Y. A. Bychkov and E. I. Rashba, *J. Phys. C: Solid State Phys.* **17**, 6039 (1984).
- <sup>34</sup> H.-J. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, *Nature Phys.* **5**, 438 (2009).
- <sup>35</sup> C.-X. Liu, X.-L. Qi, H. J. Zhang, X. Dai, Z. Fang, and S.-C. Zhang, *Phys. Rev. B* **82**, 045122 (2010).
- <sup>36</sup> Z. Li and J. P. Carbotte, *Phys. Rev. B* **89**, 085413 (2014).
- <sup>37</sup> C. J. Tabert and E. J. Nicol, *Phys. Rev. Lett.* **110**, 197402 (2013).
- <sup>38</sup> C. J. Tabert and E. J. Nicol, *Phys. Rev. B* **88**, 085434 (2013).
- <sup>39</sup> C. J. Tabert, J. P. Carbotte, and E. J. Nicol, *Phys. Rev. B* **91**, 035423 (2015).
- <sup>40</sup> V. P. Gusynin, S. G. Sharapov, and J. P. Carbotte, *J. Phys.: Condens. Matter* **19**, 026222 (2007).
- <sup>41</sup> V. P. Gusynin, S. G. Sharapov, and J. P. Carbotte, *Phys. Rev. Lett.* **96**, 256802 (2006).
- <sup>42</sup> W.-K. Tse and A. H. MacDonald, *Phys. Rev. B* **82**, 161104(R) (2010).
- <sup>43</sup> B. Scharf, V. Perebeinos, J. Fabian, and I. Žutić, *Phys. Rev. B* **88**, 125429 (2013).
- <sup>44</sup> M. Koshino and T. Ando, *Phys. Rev. B* **77**, 115313 (2008).
- <sup>45</sup> D. S. L. Abergel and V. I. Fal'ko, *Phys. Rev. B* **75**, 155430 (2007).
- <sup>46</sup> G. Tkachov and E. M. Hankiewicz, *Phys. Rev. B* **83**, 155412 (2011).
- <sup>47</sup> G. Tkachov and E. M. Hankiewicz, *Phys. Rev. B* **84**, 035405 (2011).
- <sup>48</sup> S.-Q. Shen, M. Ma, X. C. Xie, and F. C. Zhang, *Phys. Rev. Lett.* **92**, 256603 (2004).
- <sup>49</sup> S.-Q. Shen, Y.-J. Bao, M. Ma, X. C. Xie, and F. C. Zhang, *Phys. Rev. B* **71**, 155316 (2005).
- <sup>50</sup> C. J. Tabert and J. P. Carbotte, *J. Phys.: Condens. Matter* **27**, 015008 (2015).
- <sup>51</sup> Y. Suprunenko, E. V. Gorbar, S. G. Sharapov, and V. M. Loktev, *Fiz. Nizk. Temp.* **34**, 1033 (2008).
- <sup>52</sup> S. G. Sharapov, V. P. Gusynin, and H. Beck, *Phys. Rev. B* **69**, 075104 (2004).